

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>				
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To) -
4. TITLE AND SUBTITLE Non-Markovian state-dependent networks in critical loading			5a. CONTRACT NUMBER W911NF-12-1-0494	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER 611102	
6. AUTHORS Chihoon Lee, Anatolli A. Puhalskii			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Colorado State University - Ft. Collins Office of Sponsored Programs 2002 Campus Delivery Fort Collins, CO 80523 -2002			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 62402-MA-II.10	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.				
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
14. ABSTRACT We establish heavy traffic limit theorems for queue-length processes in critically loaded single class queueing networks with state-dependent arrival and service rates. A distinguishing feature of our model is non-Markovian state dependence. The limit stochastic process is a continuous-path reflected process on the nonnegative orthant. We give an application to generalised Jackson networks with state-dependent rates.				
15. SUBJECT TERMS State-dependent networks, non-Markovian networks, diffusion approximation, weak convergence				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU		
				19a. NAME OF RESPONSIBLE PERSON Chihoon Lee
				19b. TELEPHONE NUMBER 970-491-7321

Report Title

Non-Markovian state-dependent networks in critical loading

ABSTRACT

We establish heavy traffic limit theorems for queue-length processes in critically loaded single class queueing networks with state-dependent arrival and service rates. A distinguishing feature of our model is non-Markovian state dependence. The limit stochastic process is a continuous-path reflected process on the nonnegative orthant. We give an application to generalised Jackson networks with state-dependent rates.

Non-Markovian state-dependent networks in critical loading

Chihoon Lee*

Department of Statistics
Colorado State University

Anatolii A. Puhalskii†

University of Colorado Denver and
Institute for Problems in Information Transmission

January 23, 2013

Abstract

We establish heavy traffic limit theorems for queue-length processes in critically loaded single class queueing networks with state-dependent arrival and service rates. A distinguishing feature of our model is non-Markovian state dependence. The limit stochastic process is a continuous-path reflected process on the nonnegative orthant. We give an application to generalised Jackson networks with state-dependent rates.

Keywords: State-dependent networks, non-Markovian networks, diffusion approximation, weak convergence

AMS Subject Classifications: Primary 60F17; secondary 60K25, 60K30, 90B15

1 Introduction

Queueing systems with arrival and (or) service rates depending on the system's state arise in various application areas which include manufacturing, storage, service engineering, and communication and computer networks. Longer queues may lead to customers being discouraged to join the queue, or to faster processing, e.g., when human servers are involved.

*E-mail: chihoon@stat.colostate.edu

†E-mail: anatolii.puhalskii@ucdenver.edu

State-dependent features are present in congestion control protocols in communication networks, such as TCP (see [1, 3, 12, 14, 19] and references therein for more detail).

In this paper, we consider an open network of single server queues where the arrival and service rates depend on the queue lengths. The network consists of K single-server stations indexed 1 through K . Each station has an infinite capacity buffer and the customers are served according to the first-in-first-out discipline. The arrivals of customers at the stations occur both externally, from the outside and internally, from the other stations. Upon service completion at a station, a customer is either routed to another station or exits the network. Every customer entering the network eventually leaves it. A distinguishing feature of the model is non-Markovian state dependence. More specifically, the number of customers at station i , where $i = 1, 2, \dots, K$, is governed by the following equations

$$\begin{aligned} Q_i(t) &= Q_i(0) + A_i(t) + B_i(t) - D_i(t), \\ A_i(t) &= N_i^A \left(\int_0^t \lambda_i(Q(s)) ds \right), \\ B_i(t) &= \sum_{j=1}^K \Phi_{ji}(D_j(t)), \\ D_i(t) &= N_i^D \left(\int_0^t \mu_i(Q(s)) 1_{\{Q_i(s) > 0\}} ds \right), \end{aligned} \tag{1.1}$$

where $Q(s) = (Q_1(s), \dots, Q_K(s))$ denotes the vector of the queue lengths at the stations at time s . The quantities $N_i^A(t)$ and $N_i^D(t)$ represent the number of exogenous arrivals and the maximal number of customers that can be served, respectively, at station i by time t under “nominal” conditions, $\Phi_{ji}(m)$ denotes the number of customers routed from station j to station i out of the first m customers served at station j , and $\lambda_i(Q(t))$ and $\mu_i(Q(t))$ represent instantaneous exogenous arrival and service rates, respectively, for station i at time t given the queue length vector $Q(t)$. Thus, $A_i(t)$ represents the cumulative number of exogenous arrivals by time t at station i , $D_i(t)$ represents the cumulative number of departures by time t from station i , and $B_i(t)$ represents the cumulative number of customers routed to station i from the other stations by time t . The quantities $Q_i(0)$, $N_i^A(t)$, $N_i^D(t)$, and $\Phi_{ij}(m)$ are referred to as network primitives. Generalised Jackson networks is a special case of (1.1) where the N_i^A and N_i^D are renewal processes, the Φ_{ij} are Bernoulli processes, and $\lambda_i(\cdot) = \mu_i(\cdot) = 1$.

Our goal is to obtain limit theorems in critical loading for the queue length processes akin to diffusion approximation results available for generalised Jackson networks, see Reiman [18]. The results on the heavy traffic asymptotics for state-dependent rates available in the literature are mostly confined to the case of diffusion limits for Markovian models, see Yamada [20], Mandelbaum and Pats [14], and Chapter 8 of Kushner [12]. Yamada [20] and Mandelbaum and Pats [14], building on the work of Krichagina [10] who studied a Markovian closed network with state-dependent rates, considered a case of the model (1.1) where the primitive arrival and service processes are standard Poisson. Kushner [12] also includes a

treatment of that model (see Theorem 2.1 on p.318), however, their basic model is formulated in terms of the conditional distributions of the interarrival (or service) intervals (or the routing), given the “past”. Those authors obtain results in which the drift coefficients of the limit diffusion processes are state-dependent and the diffusion coefficients may be either constant or state-dependent, which is determined by the scaling used. Yamada [20] and Kushner [12] assume critical loading and obtain mostly diffusions with state-dependent diffusion coefficients, although Yamada [20] considers an example with a constant diffusion coefficient where the drift has to be linear and Theorem 2.1 on p.318 of Kushner [12] concerns the case of constant diffusion coefficients, whereas Mandelbaum and Pats [14] do not restrict their analysis to critical loading and their limits have constant diffusion coefficients. Mandelbaum and Pats [14] and Kushner [12] also allow the process of routing the customers inside the network to be state-dependent, however, their reasonings seem to be unsubstantiated, as discussed in Section 2. Section 7 of Yamada [20] is concerned with a non-Markovian case where the processes N_i^A are standard Poisson, the processes N_i^D are renewal processes, and $\mu_i(\cdot) = 1$. It is also mentioned that an extension to the case of renewal arrivals with $\lambda_i(\cdot) = 1$ and standard Poisson processes N_i^D is possible.

The main contribution of this piece of work is incorporating general arrival and service processes. This is achieved by applying an approach different from the one used by Yamada [20], Mandelbaum and Pats [14], and Kushner [12]. The proofs of those authors rely heavily on the martingale weak convergence theory. They are quite involved, on the one hand, and do not seem to be easily extendable to more general arrival and service processes, on the other hand. In our approach we, in a certain sense, return to the basics and employ ideas which have proved their worth in the set-up of generalised Jackson networks. We show that continuity considerations may produce stronger conclusions at less complexity. Our main result states that if the network primitives satisfy certain limit theorems with continuous-path limits, then the multidimensional queue-length processes, when suitably scaled and normalised, converge to a reflected continuous-path process on the nonnegative orthant. If the limits of the primitives are diffusion processes, the limit stochastic process is a reflected diffusion with state-dependent drift coefficients and constant diffusion coefficients. The scaling we use does not capture the case of state-dependent diffusion coefficients. We also give an application to generalised Jackson networks with state-dependent rates thus providing an extension of Reiman’s [18] results. In addition, we bridge certain gaps in the reasonings of Yamada [20], Mandelbaum and Pats [14], and Kushner [12]. For instance, the proofs assume functions $\lambda_i(\cdot)$ and $\mu_i(\cdot)$ are bounded and allude to a “truncation argument” for the unbounded case omitting the details. In particular, existence and uniqueness for (1.1) is not fully addressed. The key to the extension to unbounded rates is tightness of the processes in question. We give a crisp reasoning establishing that property under linear growth conditions (see Lemma 3.1), which enables us to prove both the existence and uniqueness of a solution to equations (1.1) and the limit theorem under linear growth conditions. A more detailed discussion is provided at the end of Section 2.

A different class of results on diffusion approximation concerns queueing systems modelled

on the many-server queue with a large number of servers. In such a system the service rate decreases to zero gradually with the number in the system (whereas in the model considered here it has a jump at zero, see (2.2d)), so the limit process is an unconstrained diffusion, see, Mandelbaum, Massey, and Reiman [13], Pang, Talreja, and Whitt [15], and references therein. We do not consider those set-ups in this paper.

The exposition is organised as follows. In the next section, we state and discuss our main result. The proof is provided in Section 3. In Section 4, an application to state-dependent generalised Jackson networks is presented. The appendix contains a proof of the pathwise queue-length construction underlying the definition of the model.

Some notational conventions are in order. All vectors are understood as column vectors, $|x|$ denotes the Euclidean length of a vector x , its components are denoted by x_i , unless mentioned otherwise, superscript T is used to denote the transpose, 1_A stands for the indicator function of an event A , δ_{ij} represents Kronecker's delta, $[a]$ denotes the integer part of a real number a , \mathbb{Z}_+ denotes the set of whole numbers, and S is used to denote the K -dimensional non-negative orthant \mathbb{R}_+^K . We use $\mathbb{D}([0, \infty), \mathbb{R}^\ell)$ to represent the Skorohod space of right continuous functions with left hand limits which is endowed with the Skorohod topology, \Rightarrow represents convergence in distribution of random elements with values in an appropriate metric space, see Billingsley [2], Ethier and Kurtz [6] for more information. We also recall that a sequence V^n of stochastic processes with trajectories in a Skorohod space is said to be \mathbb{C} -tight if the sequence of the laws of the V^n is tight, and if all limit points of the sequence of the laws of the V^n are laws of continuous-path processes (see, e.g., Definition 3.25 and Proposition 3.26 in Chapter VI of Jacod and Shiryaev [9]).

2 The main result

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space where all random variables considered in this paper are assumed to be defined. We consider a sequence of networks indexed by n with a similar structure as the one described in the Introduction. For the n -th network and for $i \in \mathbb{K}$, where $\mathbb{K} = \{1, 2, \dots, K\}$, let $A_i^n(t)$ represent the cumulative number of customers that arrive at station i from outside the network during the time interval $[0, t]$, and let $D_i^n(t)$ represent the cumulative number of customers that are served at station i for the first t units of busy time of that station. Let $\mathcal{J} \subseteq \mathbb{K}$ represent the set of stations with actual arrivals so that $A_i^n(t) = 0$ if $i \notin \mathcal{J}$. We call $A^n = (A_i^n, i \in \mathbb{K})$ and $D^n = (D_i^n, i \in \mathbb{K})$, where $A_i^n = (A_i^n(t), t \geq 0)$ and $D_i^n = (D_i^n(t), t \geq 0)$, the arrival process and service process for the n -th network, respectively. We associate with the stations of the network the processes $\Phi_i^n = (\Phi_{ij}^n, j \in \mathbb{K})$, $i \in \mathbb{K}$, where $\Phi_{ij}^n = (\Phi_{ij}^n(m), m = 1, 2, \dots)$, and $\Phi_{ij}^n(m)$ denotes the cumulative number of customers among the first m customers that depart station i which go directly to station j . The process $\Phi^n = (\Phi_{ij}^n, i, j \in \mathbb{K})$ is referred to as the routing process. We consider the processes A_i^n, D_i^n , and Φ_i^n as random elements of the respective Skorohod spaces $\mathbb{D}([0, \infty), \mathbb{R}), \mathbb{D}([0, \infty), \mathbb{R})$, and $\mathbb{D}([0, \infty), \mathbb{R}^K)$; accordingly, A^n, D^n , and

Φ^n are regarded as random elements of $\mathbb{D}([0, \infty), \mathbb{R}^K)$, $\mathbb{D}([0, \infty), \mathbb{R}^K)$, and $\mathbb{D}([0, \infty), \mathbb{R}^{K \times K})$, respectively.

Let λ_i^n and μ_i^n , where $i \in \mathbb{K}$, be Borel functions mapping S to \mathbb{R}_+ , with $\lambda_i^n(x) = 0$ if $i \notin \mathcal{J}$, and let $\lambda^n = (\lambda_1^n, \dots, \lambda_K^n)$ and $\mu^n = (\mu_1^n, \dots, \mu_K^n)$. These functions have the meaning of state-dependent arrival and service rates. Let $N_i^{A,n} = (N_i^{A,n}(t), t \geq 0)$ and $N_i^{D,n} = (N_i^{D,n}(t), t \geq 0)$ represent nondecreasing \mathbb{Z}_+ -valued processes with trajectories in $\mathbb{D}([0, \infty), \mathbb{R})$ and with $N_i^{A,n}(0) = N_i^{D,n}(0) = 0$. We define $N_i^{A,n}(t) = \lfloor t \rfloor$ if $i \notin \mathcal{J}$. (The latter is but a convenient convention. Since $\lambda_i^n(x) = 0$ if $i \notin \mathcal{J}$, the process $N_i^{A,n}$ is immaterial, as the equations below show.) The state of the network at time t is represented by $Q^n(t) = (Q_1^n(t), \dots, Q_K^n(t))$, where $Q_i^n(t)$ represents the number of customers at station i at time t . It is assumed to satisfy a.s. the equations:

$$Q_i^n(t) = Q_i^n(0) + A_i^n(t) + B_i^n(t) - D_i^n(t), \quad (2.2a)$$

$$A_i^n(t) = N_i^{A,n} \left(\int_0^t \lambda_i^n(Q^n(s)) ds \right), \quad (2.2b)$$

$$B_i^n(t) = \sum_{j=1}^K \Phi_{ji}^n(D_j^n(t)), \quad (2.2c)$$

$$D_i^n(t) = N_i^{D,n} \left(\int_0^t \mu_i^n(Q^n(s)) 1_{\{Q_i^n(s) > 0\}} ds \right), \quad (2.2d)$$

where $t \geq 0$ and $i \in \mathbb{K}$. As above, $Q_i^n(0) \in \mathbb{Z}_+$ is the initial queue length at station i ; $A_i^n(t)$, $B_i^n(t)$, and $D_i^n(t)$ represent the cumulative number of exogenous arrivals at station i during the time interval $[0, t]$, the cumulative number of endogenous arrivals at station i during the time interval $[0, t]$, and the cumulative number of departures from station i during the time interval $[0, t]$, respectively.

Let $P = (p_{ij}, i, j \in \mathbb{K})$ be a substochastic matrix, $R = I - P^T$, and $p_i = (p_{ij}, j \in \mathbb{K})$. We denote

$$\begin{aligned} \bar{Q}^n(0) &= \frac{Q^n(0)}{\sqrt{n}}, & \bar{N}_i^{A,n}(t) &= \frac{N_i^{A,n}(nt) - nt}{\sqrt{n}}, \\ \bar{N}_i^{D,n}(t) &= \frac{N_i^{D,n}(nt) - nt}{\sqrt{n}}, & \bar{\Phi}_i^n(t) &= \frac{\Phi_i^n(\lfloor nt \rfloor) - p_i nt}{\sqrt{n}}, \\ \bar{N}_i^{A,n} &= (\bar{N}_i^{A,n}(t), t \geq 0), & \bar{N}^{A,n} &= (\bar{N}_i^{A,n}, i \in \mathbb{K}), \\ \bar{N}_i^{D,n} &= (\bar{N}_i^{D,n}(t), t \geq 0), & \bar{N}^{D,n} &= (\bar{N}_i^{D,n}, i \in \mathbb{K}), \\ \bar{\Phi}_i^n &= (\bar{\Phi}_i^n(t), t \geq 0), & \bar{\Phi}^n &= (\bar{\Phi}_i^n, i \in \mathbb{K}). \end{aligned}$$

We will need the following conditions.

(A0) For each $n \in \mathbb{N}$ and each $i \in \mathcal{J}$, $\limsup_{t \rightarrow \infty} N_i^{A,n}(t)/t < \infty$ a.s.

(A1) The spectral radius of matrix P is strictly less than 1.

(A2) For each $i \in \mathbb{K}$,

$$\sup_{n \in \mathbb{N}} \sup_{x \in S} \frac{\lambda_i^n(nx) + \mu_i^n(nx)}{n(1 + |x|)} < \infty.$$

(A3) There exist continuous functions $\lambda_i(x)$ and $\mu_i(x)$ such that

$$\frac{\lambda_i^n(nx)}{n} \rightarrow \lambda_i(x), \quad \frac{\mu_i^n(nx)}{n} \rightarrow \mu_i(x)$$

uniformly on compact subsets of S , as $n \rightarrow \infty$. Furthermore, for $x \in S$,

$$\lambda(x) - R\mu(x) = 0.$$

(A4) There exists a Lipschitz-continuous function $a(x)$ such that

$$\frac{1}{\sqrt{n}} (\lambda^n(\sqrt{n}x) - R\mu^n(\sqrt{n}x)) \rightarrow a(x)$$

as $n \rightarrow \infty$ uniformly on compact subsets of S .

(A5) As $n \rightarrow \infty$,

$$(\overline{Q}^n(0), \overline{N}^{A,n}, \overline{N}^{D,n}, \overline{\Phi}^n) \Rightarrow (X_0, W^A, W^D, W^\Phi)$$

where X_0 is a random K -vector, W^A , W^D , and W^Φ are continuous-path stochastic processes with trajectories in respective spaces $\mathbb{D}([0, \infty), \mathbb{R}^K)$, $\mathbb{D}([0, \infty), \mathbb{R}^K)$, and $\mathbb{D}([0, \infty), \mathbb{R}^{K \times K})$.

Lemma 2.1. *Let condition (A0) hold and $\max_{i \in \mathbb{K}} \sup_{x \in S} (\lambda_i^n(x) + \mu_i^n(x)) / (1 + |x|) < \infty$. Then equations (2.2a)–(2.2d) admit a unique strong solution Q^n , which is a \mathbb{Z}_+^K -valued stochastic process.*

The proof is provided in the appendix. In order to state the main result, we have to recall standard properties of the Skorohod map.

Definition 2.2. *Let $\psi \in \mathbb{D}([0, \infty), \mathbb{R}^K)$ be given with $\psi(0) \in S$. Then the pair $(\phi, \eta) \in \mathbb{D}([0, \infty), \mathbb{R}^K) \times \mathbb{D}([0, \infty), \mathbb{R}^K)$ solves the Skorohod problem for ψ with respect to S and R if the following hold:*

- (i) $\phi(t) = \psi(t) + R\eta(t) \in S$, for all $t \geq 0$;
- (ii) for $i \in \mathbb{K}$, (a) $\eta_i(0) = 0$, (b) η_i is non-decreasing, and (c) η_i can increase only when ϕ is on the i^{th} face of S , that is, $\int_0^\infty 1_{\{\phi_i(s) \neq 0\}} d\eta_i(s) = 0$.

Let $\mathbb{D}_S([0, \infty), \mathbb{R}^K) = \{\psi \in \mathbb{D}([0, \infty), \mathbb{R}^K) : \psi(0) \in S\}$. If the Skorohod problem has a unique solution on a domain $D \subset \mathbb{D}_S([0, \infty), \mathbb{R}^K)$, we define the Skorohod map Γ on D by $\Gamma(\psi) = \phi$. The following result (see Harrison and Reiman [7] and also Dupuis and Ishii [4]) yields the regularity of the Skorohod map and is a consequence of Assumption (A1).

Proposition 2.3. *The Skorohod map Γ is well defined on $\mathbb{D}_S([0, \infty), \mathbb{R}^K)$ and is Lipschitz continuous in the following sense: There exists a constant $L > 0$ such that for all $T > 0$ and $\psi_1, \psi_2 \in \mathbb{D}_S([0, \infty), \mathbb{R}^K)$,*

$$\sup_{t \in [0, T]} |\Gamma(\psi_1)(t) - \Gamma(\psi_2)(t)| \leq L \sup_{t \in [0, T]} |\psi_1(t) - \psi_2(t)|.$$

Consequently, both ϕ and η are continuous functions of ψ .

The Lipschitz continuity of the Skorohod map and of the function $a(x)$ imply that the equation

$$X(t) = \Gamma\left(X_0 + \int_0^t a(X(s))ds + M(\cdot)\right)(t), \quad (2.3)$$

where

$$M_i(t) = W_i^A(\lambda_i(0)t) + \sum_{j=1}^K W_{ji}^\Phi(\mu_j(0)t) - \sum_{j=1}^K (\delta_{ij} - p_{ji})W_j^D(\mu_j(0)t), \quad (2.4)$$

has a unique strong solution. For $t \geq 0$ and $i \in \mathbb{K}$, let $X_i^n(t) = Q_i^n(t)/\sqrt{n}$. We also define $X = (X(t), t \geq 0)$ and $X^n = ((X_i^n(t), i = 1, 2, \dots, K), t \geq 0)$.

Theorem 2.4. *Let conditions (A0)–(A5) hold. Then $X^n \Rightarrow X$, as $n \rightarrow \infty$.*

The proof is given in the next section.

We now discuss our results as well as those of Yamada [20] and Mandelbaum and Pats [14]. Condition (A0) is needed to ensure the existence of a unique strong solution to the system of equations (2.2a)–(2.2d), see Lemma 2.1. It is certainly fulfilled if $N_i^{A,n}$ is a renewal process and is almost a consequence of condition (A5) in that the latter implies that $\lim_{n \rightarrow \infty} N_i^{A,n}(nt)/(nt) = 1$ in probability. Part (A1) is essentially an assumption that the network is open and underlies the existence of a regular Skorohod map associated with the network data asserted in Proposition 2.3. The linear growth condition (A2) is the same as condition (A2) in Mandelbaum and Pats [14]. The requirement $\lambda(x) = R\mu(x)$ in (A3) together with condition (A4) defines a *critically loaded heavy traffic* regime. Condition (A5) is the assumption on the primitives. The components of W^A corresponding to $i \notin \mathcal{J}$ vanish. Conditions (A2)–(A4) are fulfilled if the following expansions hold:

$$\lambda^n(x) = n\lambda_1(x/n) + \sqrt{n}\lambda_2(x/\sqrt{n}) \text{ and } \mu^n(x) = n\mu_1(x/n) + \sqrt{n}\mu_2(x/\sqrt{n}), \quad (2.5)$$

where λ_1 and μ_1 are continuous functions satisfying the linear-growth condition such that $\lambda_1(x) = R\mu_1(x)$, and λ_2 and μ_2 are bounded Lipschitz-continuous. If the above functions are constant, then one obtains the standard critical loading condition that $(\lambda^n - R\mu^n)/\sqrt{n} \rightarrow \lambda_2 - \mu_2$ as $n \rightarrow \infty$, cf. Reiman [18].

Most of the results on diffusion approximation in critical loading (see, e.g., Harrison and Reiman [7], Kushner [12]) formulate the heavy traffic condition in terms of rates that are

$\mathcal{O}(1)$ and then consider scaled processes $Q^n(nt)/\sqrt{n}$. In the scaling considered here, as in Yamada [20] and Mandelbaum and Pats [14], the time parameter is left unchanged and the factor of n is absorbed in the arrival and service rates. This is more convenient notationally, however, in the application to generalised Jackson networks in Section 4 we work with the conventional scaling. It may be instructive to note, though, that if one looked for limits for processes $Q^n(nt)/\sqrt{n}$, then the analogues of expansions (2.5) would be

$$\lambda^n(x) = \lambda_1(x/n) + (1/\sqrt{n})\lambda_2(x/\sqrt{n}) \text{ and } \mu^n(x) = \mu_1(x/n) + (1/\sqrt{n})\mu_2(x/\sqrt{n}),$$

whereas the assumptions of Yamada [20] would amount to the expansions

$$\lambda^n(x) = \lambda_1(x/\sqrt{n}) + (1/\sqrt{n})\lambda_2(x/\sqrt{n}) \text{ and } \mu^n(x) = \mu_1(x/\sqrt{n}) + (1/\sqrt{n})\mu_2(x/\sqrt{n}).$$

Theorems 1 and 2 in Yamada [20] obtain diffusion processes with state-dependent drift and diffusion coefficients as the limits. Theorem 1 concerns the Markovian model. It is required that there be at least one nonzero external arrival process. The arrival and service rates at a station may depend on the queue length at that station only. Theorem 2 concerns a Jackson network with external arrival processes being Poisson processes with state-dependent rates. In the proof of Theorem 2, the process $d_n^j(t)$ claimed to be a locally square integrable martingale on p.980 does not seem to be so, anyway, no justification is provided.

The model of Yamada [20] is defined by postulating certain martingale properties of the arrival, service, and customer transfer processes. No justification of this model in terms of the primitive processes is provided, nor is the issue of the assumptions being self-consistent addressed. This gap is filled in by Mandelbaum and Pats [14], although the authors admit the proof is missing technical detail, see p.623 in Mandelbaum and Pats [14]. On the other hand, Mandelbaum and Pats [14] produce few details with regard to the existence and uniqueness of a solution to (1.1) under linear growth conditions on the rates. In particular, it is not explained how Theorem 2.1 in Kurtz [11] enables one to establish Proposition 13.4, nor is it spelled out how Proposition 13.4 furnishes the proof of existence and uniqueness. The proof of Theorem 2.1 in Kurtz [11] seems to have gaps too. Mandelbaum and Pats [14] and Kushner [12] allow the routing matrix to be state-dependent. Mandelbaum and Pats [14] appeal to Theorem 5.1 and Corollary 5.2 in Dupuis and Ishii [5] to substantiate the existence and uniqueness for the Skorohod problem, however, those results are proved for bounded domains, so they do not apply. The authors' attempt on p.628 to recast the reflection problem as a time-dependent reflection is unconvincing. Kushner [12], in their proofs of Theorem 1.1 on p.309 and Theorem 2.1 on p.318, relies on their Theorem 5.1 on p.123 and Theorem 5.2 on p.124 which in turn are based on Theorem 2.2 in Dupuis and Ishii [4], however, those results pertain to reflection directions which are constant on the faces, so they do not apply to state-dependent reflection directions.

Kushner [12] does not address the issue of the model being well defined either. Nor are we convinced by the substantiation of martingale properties claimed to hold on p.310. Besides, the hypotheses of Theorem 1.1 on p.309 and Theorem 2.1 on p.318 of Kushner [12]

are missing the condition of the drift and diffusion coefficients being Lipschitz continuous. On the other hand, the condition that the first moments of the initial queue lengths be finite assumed by Mandelbaum and Pats [14] can be done away with.

3 Proof of Theorem 2.4

We assume conditions (A0)–(A5) throughout this section. We introduce the “centered” processes as follows: For $i \in \mathbb{K}$ and $t \geq 0$,

$$M_i^n(t) = M_i^{A,n}(t) + M_i^{B,n}(t) - M_i^{D,n}(t), \quad (3.6a)$$

where

$$M_i^{A,n}(t) = N_i^{A,n} \left(\int_0^t \lambda_i^n(Q^n(s)) ds \right) - \int_0^t \lambda_i^n(Q^n(s)) ds, \quad (3.6b)$$

$$M_i^{B,n}(t) = \sum_{j=1}^K (\Phi_{ji}^n(D_j^n(t)) - p_{ji} D_j^n(t)), \quad (3.6c)$$

and

$$\begin{aligned} M_i^{D,n}(t) = & N_i^{D,n} \left(\int_0^t \mu_i^n(Q^n(s)) 1_{\{Q_i^n(s) > 0\}} ds \right) - \int_0^t \mu_i^n(Q^n(s)) 1_{\{Q_i^n(s) > 0\}} ds \\ & + \sum_{j=1}^K p_{ji} \left(N_j^{D,n} \left(\int_0^t \mu_j^n(Q^n(s)) 1_{\{Q_j^n(s) > 0\}} ds \right) - \int_0^t \mu_j^n(Q^n(s)) 1_{\{Q_j^n(s) > 0\}} ds \right). \end{aligned} \quad (3.6d)$$

We can rewrite the evolution (2.2a) as

$$Q_i^n(t) = Q_i^n(0) + \int_0^t \left[\lambda_i^n(Q^n(s)) + \sum_{j=1}^K p_{ji} \mu_j^n(Q^n(s)) - \mu_i^n(Q^n(s)) \right] ds + M_i^n(t) + [RY^n(t)]_i,$$

where $Y^n(t) = (Y_i^n(t), i \in \mathbb{K})$ and

$$Y_i^n(t) = \int_0^t 1_{\{Q_i^n(s)=0\}} \mu_i^n(Q^n(s)) ds, \quad i \in \mathbb{K}, \quad (3.7)$$

Note that $(Y_i^n(t), t \geq 0)$ is a continuous-path non-decreasing process with $Y_i^n(0) = 0$, which increases only when $Q_i^n(t) = 0$, i.e., $\int_0^\infty 1_{\{Q_i^n(t) \neq 0\}} dY_i^n(t) = 0$ a.s. Let

$$a^n(x) = \lambda^n(x) - R\mu^n(x). \quad (3.8)$$

Then the state evolution can be expressed succinctly by the following vector equation:

$$Q^n(t) = Q^n(0) + \int_0^t a^n(Q^n(s))ds + M^n(t) + RY^n(t), \quad t \geq 0. \quad (3.9)$$

It can also be described in terms of the Skorohod map:

$$Q^n(t) = \Gamma \left(Q^n(0) + \int_0^\cdot a^n(Q^n(s))ds + M^n(\cdot) \right) (t), \quad \text{for } t \geq 0. \quad (3.10)$$

The following tightness result is essential.

Lemma 3.1. *The sequence of processes $(M^n(t)/\sqrt{n}, t \geq 0)$ is \mathbb{C} -tight.*

Proof. By (2.2a) – (2.2d),

$$\sum_{i=1}^K Q_i^n(t) \leq \sum_{i=1}^K Q_i^n(0) + \sum_{i=1}^K A_i^n(t) = \sum_{i=1}^K Q_i^n(0) + \sum_{i=1}^K N_i^{A,n} \left(\int_0^t \lambda_i^n(Q^n(s))ds \right).$$

Therefore, for suitable $H > 0$, on recalling (A2) and denoting $Z_i^n(t) = Q_i^n(t)/n$,

$$\begin{aligned} \sum_{i=1}^K Z_i^n(t) &\leq \sum_{i=1}^K Z_i^n(0) + \sum_{i=1}^K \sup_{y \geq n} \frac{1}{y} N_i^{A,n}(y) \left(1 + \int_0^t \frac{1}{n} \lambda_i^n(nZ^n(s))ds \right) \\ &\leq \sum_{i=1}^K Z_i^n(0) + \sum_{i=1}^K \sup_{y \geq n} \frac{1}{y} N_i^{A,n}(y) \left(1 + H \int_0^t (1 + \sum_{i=1}^K Z_i^n(s)) ds \right). \end{aligned}$$

By Gronwall's inequality (cf. p.498 in Ethier and Kurtz [6]),

$$\sum_{i=1}^K Z_i^n(t) \leq \left(\sum_{i=1}^K Z_i^n(0) + \sum_{i=1}^K \sup_{y \geq n} \frac{1}{y} N_i^{A,n}(y)(1 + Ht) \right) \exp \left(H \sum_{i=1}^K \sup_{y \geq n} \frac{1}{y} N_i^{A,n}(y)t \right).$$

By (A5), $N_i^{A,n}(y)/y \rightarrow 1$ in probability as $y \rightarrow \infty$ and $n \rightarrow \infty$ and $\sum_{i=1}^K Z_i^n(0) \rightarrow 0$ in probability as $n \rightarrow \infty$. Therefore,

$$\lim_{r \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P} \left(\sup_{s \leq t} \sum_{i=1}^K Z_i^n(s) > r \right) = 0. \quad (3.11)$$

It follows by (A2) that

$$\lim_{r \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P} \left(\int_0^t \left(\frac{1}{n} \lambda_i^n(Q_i^n(s)) + \frac{1}{n} \mu_i^n(Q_i^n(s)) \right) ds > r \right) = 0 \quad (3.12)$$

and that, for $\delta > 0$, $\epsilon > 0$, $T > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T]} \int_t^{t+\delta} \left(\frac{1}{n} \lambda_i^n(Q_i^n(s)) + \frac{1}{n} \mu_i^n(Q_i^n(s)) \right) ds > \epsilon \right) = 0. \quad (3.13)$$

We have that, for $\gamma > 0$, $\delta > 0$, $\epsilon > 0$, $T > 0$, and $r > 0$,

$$\begin{aligned} \mathbf{P} \left(\sup_{\substack{s, t \in [0, T]: \\ |s-t| \leq \delta}} \left| \frac{1}{\sqrt{n}} M_i^{A,n}(t) - \frac{1}{\sqrt{n}} M_i^{A,n}(s) \right| > \gamma \right) &\leq \mathbf{P} \left(\int_0^T \frac{1}{n} \lambda_i^n(Q_i^n(s)) ds > r \right) \\ &+ \mathbf{P} \left(\sup_{t \in [0, T]} \int_t^{t+\delta} \frac{1}{n} \lambda_i^n(Q_i^n(s)) ds > \epsilon \right) + \mathbf{P} \left(\sup_{\substack{s, t \in [0, r]: \\ |s-t| \leq \epsilon}} |\bar{N}_i^{A,n}(t) - \bar{N}_i^{A,n}(s)| > \gamma \right). \end{aligned}$$

By (A5), (3.12), and (3.13),

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbf{P} \left(\sup_{\substack{s, t \in [0, T]: \\ |s-t| \leq \delta}} \left| \frac{1}{\sqrt{n}} M_i^{A,n}(t) - \frac{1}{\sqrt{n}} M_i^{A,n}(s) \right| > \gamma \right) = 0.$$

Hence, the sequences of processes $(M_i^{A,n}(t)/\sqrt{n}, t \geq 0)$ are \mathbb{C} -tight. A similar argument shows that the sequences of processes $(M_i^{D,n}(t)/\sqrt{n}, t \geq 0)$ and $(M_i^{\Phi,n}(t)/\sqrt{n}, t \geq 0)$ are \mathbb{C} -tight, so the sequence of processes $(M^n(t)/\sqrt{n}, t \geq 0)$ is \mathbb{C} -tight. \blacksquare

Next, we identify the limit points of $\bar{M}^n = (M^n(t)/\sqrt{n}, t \geq 0)$.

Lemma 3.2. *The sequence of processes \bar{M}^n converges in distribution, as $n \rightarrow \infty$, to M .*

Proof. From Lemma 3.1, $M^n(t)/n \rightarrow 0$ in probability uniformly over bounded intervals. By (A2) and (3.8), for some H' , for all n and x , $|a^n(nx)| \leq H'n(1 + |x|)$. By (3.11), the sequence of processes $(\int_0^t (1/n) a^n(Q^n(s)) ds, t \geq 0)$ is \mathbb{C} -tight. By the fact that $M^n(t)/n \rightarrow 0$ in probability uniformly on bounded intervals, Prohorov's theorem, and the continuity of the Skorohod map, the sequence of processes $(Q^n(t)/n, t \geq 0)$ is \mathbb{C} -tight and every its limit in distribution $(q(t), t \geq 0)$ satisfies the equation

$$q(t) = \Gamma \left(\int_0^t (\lambda(q(s)) - R\mu(q(s))) ds \right) (t).$$

Since by (A3), $\lambda(x) - R\mu(x) = 0$, we must have that $q(t) = 0$, which implies that the sequence $Q_i^n(t)/n$ tends to zero as $n \rightarrow \infty$ in probability uniformly on bounded intervals. By condition (A3) and (3.8), $\int_0^t (1/n) a^n(Q^n(s)) ds \rightarrow 0$ in probability. Since by (3.9) Y^n is expressed as a continuous function of $(Q^n(0) + \int_0^t a^n(Q^n(s)) ds + M^n(t), t \geq 0)$, we have that $Y^n(t)/n \rightarrow 0$ in probability uniformly over bounded intervals, so by (3.7), for $i \in \mathbb{K}$,

$$\frac{1}{n} \int_0^t \mu_i^n(Q^n(s)) 1_{\{Q_i^n(s)=0\}} ds \rightarrow 0 \text{ in probability as } n \rightarrow \infty. \quad (3.14)$$

We also have by (A3) that

$$\frac{1}{n} \int_0^t \lambda_i^n(Q^n(s)) ds \rightarrow \lambda_i(0)t \text{ in probability as } n \rightarrow \infty \quad (3.15a)$$

and

$$\frac{1}{n} \int_0^t \mu_i^n(Q^n(s)) ds \rightarrow \mu_i(0)t \text{ in probability as } n \rightarrow \infty. \quad (3.15b)$$

Since by (A5), $N_i^{D,n}(nt)/n \rightarrow t$ in probability as $n \rightarrow \infty$, by (2.2d), (A4), (3.14), and (3.15b),

$$\frac{D_i^n(t)}{n} \rightarrow \mu_i(0)t \text{ in probability as } n \rightarrow \infty. \quad (3.16)$$

The convergences in (A5), (3.15a), (3.15b), and (3.16) imply if one recalls the definitions in (3.6b), (3.6c), and (3.6d) that the $(M^{A,n}/\sqrt{n}, M^{B,n}/\sqrt{n}, M^{D,n}/\sqrt{n})$ converge in distribution to (M^A, M^B, M^D) , where $M_i^A(t) = W_i^A(\lambda_i(0)t)$, $M_i^B(t) = \sum_{j=1}^K W_{ji}^B(\mu_j(0)t)$, $M_i^D(t) = \sum_{j=1}^K (\delta_{ij} - p_{ji}) W_j^D(\mu_j(0)t)$, so, by (3.6a) and (2.4), the \bar{M}^n converge in distribution to \bar{M} . ■

Proof of Theorem 2.4. We note that by (3.10),

$$X^n(t) = \Gamma \left(X^n(0) + \int_0^t \frac{1}{\sqrt{n}} a^n(\sqrt{n}X^n(s)) ds + \bar{M}^n(\cdot) \right) (t), \quad \text{for } t \geq 0. \quad (3.17)$$

By the Lipschitz continuity of the Skorohod map, (3.8), and (A2), for $T > 0$ and suitable $H > 0$,

$$\begin{aligned} \sup_{t \in [0, T]} |X^n(t)| &\leq |X^n(0)| + L \int_0^t \frac{1}{\sqrt{n}} |a^n(\sqrt{n}X^n(s))| ds + \frac{1}{\sqrt{n}} \sup_{t \in [0, T]} |M^n(t)| \\ &\leq |X^n(0)| + LH \int_0^t (1 + |X^n(s)|) ds + \frac{1}{\sqrt{n}} \sup_{t \in [0, T]} |M^n(t)|. \end{aligned}$$

Gronwall's inequality, the convergence of the $X^n(0)$, and Lemma 3.2 yield

$$\lim_{r \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T]} |X^n(t)| > r \right) = 0,$$

so, by (3.8) and (A4), the sequence of processes $(\int_0^t a^n(\sqrt{n}X^n(s))/\sqrt{n} ds, t \geq 0)$ is \mathbb{C} -tight.

By (3.17), the convergence of the $X^n(0)$, Lemma 3.2, (A4), Prohorov's theorem, and the continuity of the Skorohod map, the sequence of processes $(X^n(t), t \geq 0)$ is \mathbb{C} -tight and every limit point $(\tilde{X}(t), t \geq 0)$ for convergence in distribution satisfies the equation

$$\tilde{X}(t) = \Gamma \left(X(0) + \int_0^t a(\tilde{X}(s)) ds + M(\cdot) \right) (t), \quad \text{for } t \geq 0.$$

The uniqueness of a solution to the Skorohod problem implies that $\tilde{X}(t) = X(t)$. ■

4 Generalised Jackson networks with state-dependent rates

In this section, we consider an application to generalised Jackson networks in conventional scaling. Suppose as given mutually independent sequences of i.i.d. nonnegative random variables $\{u_j^i(n), i \geq 1\}$, $\{v_k^i(n), i \geq 1\}$ for $j \in \mathcal{J} \subseteq \mathbb{K}$ and $k \in \mathbb{K}$. For the n th network, the random variable $u_j^i(n)$ represents the i th exogenous interarrival time at station j , while $v_k^i(n)$ is the i th service time at station k . The quantities p_{ij} represent the probabilities of a customer leaving station i being routed directly to station j , which are held constant. The routing decisions, interarrival and service times, and the initial queue length vector are mutually independent.

We define

$$\begin{aligned} \mu_k^n &= (\mathbf{E}[v_k^1(n)])^{-1} > 0, & s_k^n &= \mathbf{Var}(v_k^1(n)) \geq 0, & k &\in \mathbb{K}, & \text{and} \\ \lambda_j^n &= (\mathbf{E}[u_j^1(n)])^{-1} > 0, & a_j^n &= \mathbf{Var}(u_j^1(n)) \geq 0, & j &\in \mathcal{J}, \end{aligned}$$

with all of these terms assumed finite and the set \mathcal{J} nonempty. It is convenient to let $\lambda_j^n = 1$ and $a_j^n = 0$ for $j \notin \mathcal{J}$.

Let $\hat{N}_j^{\hat{A},n}(t) = \max\{i' : \sum_{i=1}^{i'} u_j^i(n) \leq t\}$ for $j \in \mathcal{J}$ and $\hat{N}_k^{\hat{D},n}(t) = \max\{i' : \sum_{i=1}^{i'} v_k^i(n) \leq t\}$ for $k \in \mathbb{K}$. We may interpret the process $(\hat{N}_j^{\hat{A},n}(t), t \geq 0)$ as a nominal arrival process and the random variables $v_k^i(n)$ as the amounts of work needed to serve the customers. Suppose that arrivals are speeded up (or delayed) by a function $\hat{\lambda}_i^n(x)$, where $i \in \mathcal{J}$, and the service is performed at rate $\hat{\mu}_k^n(x)$, where $k \in \mathbb{K}$, when the queue length vector is x . As in Section 3, we let $\hat{N}_i^{\hat{A},n}(t) = \lfloor t \rfloor$ and $\hat{\lambda}_i^n(x) = 0$ for $i \notin \mathcal{J}$. In analogy with (2.2a)-(2.2d) the queue lengths at the stations at time t , which we denote by $\hat{Q}_i^n(t)$, are assumed to satisfy

the equations

$$\begin{aligned}\hat{Q}_i^n(t) &= \hat{Q}_i^n(0) + \hat{A}_i^n(t) + \hat{B}_i^n(t) - \hat{D}_i^n(t), \\ \hat{A}_i^n(t) &= \hat{N}_i^{\hat{A},n} \left(\int_0^t \hat{\lambda}_i^n(\hat{Q}^n(s)) ds \right), \\ \hat{B}_i^n(t) &= \sum_{j=1}^K \hat{\Phi}_{ji}^n(\hat{D}_j^n(t)), \\ \hat{D}_i^n(t) &= \hat{N}_i^{\hat{D},n} \left(\int_0^t \hat{\mu}_i^n(\hat{Q}^n(s)) 1_{\{\hat{Q}_i^n(s) > 0\}} ds \right),\end{aligned}$$

where

$$\hat{\Phi}_{ji}^n(m) = \sum_{l=1}^m \chi_{ji}^n(l),$$

with $\{(\chi_{ji}^n(l), i = 1, 2, \dots, K), l = 1, 2, \dots\}$ being indicator random variables which are mutually independent for different j and l and are such that $\mathbf{P}(\chi_{ji}^n(l) = 1) = p_{ji}$.

If we introduce the random variables $Q_i^n(t) = \hat{Q}_i^n(nt)$, $A_i^n(t) = \hat{A}_i^n(nt)$, $B_i^n(t) = \hat{B}_i^n(nt)$, $D_i^n(t) = \hat{D}_i^n(nt)$, $N_i^{A,n}(t) = \hat{N}_i^{\hat{A},n}(t/\lambda_i^n)$, $N_i^{D,n}(t) = \hat{N}_i^{\hat{D},n}(t/\mu_i^n)$, and $\Phi_{ji}^n(m) = \hat{\Phi}_{ji}^n(m)$, and functions $\lambda_i^n(x) = n\lambda_i^n\hat{\lambda}_i^n(x)$ and $\mu_i^n(x) = n\mu_i^n\hat{\mu}_i^n(x)$, then we can see that they satisfy equations (2.2a)–(2.2d). Condition (A0) holds as $N_i^{A,n}(t)/t \rightarrow 1$ and $N_i^{D,n}(t)/t \rightarrow 1$ a.s. as $t \rightarrow \infty$.

If we also assume that $\hat{Q}^n(0)/\sqrt{n} \Rightarrow X_0$, that, for $k \in \mathbb{K}$ and $j \in \mathcal{J}$,

$$\begin{aligned}\mu_k^n &\rightarrow \mu_k, & s_k^n &\rightarrow s_k, \\ \lambda_j^n &\rightarrow \lambda_j, & a_j^n &\rightarrow a_j,\end{aligned}$$

as $n \rightarrow \infty$, and that

$$\max_{k \in \mathbb{K}} \sup_{n \geq 1} \mathbf{E}(v_k^1(n))^{2+\epsilon} + \max_{j \in \mathcal{J}} \sup_{n \geq 1} \mathbf{E}(u_j^1(n))^{2+\epsilon} < \infty \quad \text{for some } \epsilon > 0,$$

then condition (A5) holds with $W_j^A = \sqrt{a_j} \lambda_j B_j^A$ for $j \in \mathcal{J}$, $W_j^A(t) = 0$ for $j \notin \mathcal{J}$, and $W_k^D = \sqrt{s_k} \mu_k B_k^D$ for $k \in \mathbb{K}$, where B_j^A and B_k^D are independent standard Brownian motions, with Φ_i being a K -dimensional Brownian motion with covariance matrix $\mathbf{E}\Phi_{ik}(t)\Phi_{ij}(t) = (p_{ij}\delta_{jk} - p_{ij}p_{ik})t$, and with processes B_j^A , B_k^D , and Φ_i being mutually independent.

Let us assume that the following versions of conditions (A2)–(A4) hold:

(A2) For each $i \in \mathbb{K}$,

$$\sup_{n \in \mathbb{N}} \sup_{x \in S} \left(\frac{\hat{\lambda}_i^n(nx)}{1 + |x|} + \frac{\hat{\mu}_i^n(nx)}{1 + |x|} \right) < \infty,$$

(A3) There exist continuous functions $\hat{\lambda}_i(x)$ and $\hat{\mu}_i(x)$ such that

$$\hat{\lambda}_i^n(nx) \rightarrow \hat{\lambda}_i(x), \quad \hat{\mu}_i^n(nx) \rightarrow \hat{\mu}_i(x)$$

uniformly on compact subsets of S , as $n \rightarrow \infty$. Furthermore, for $x \in S$,

$$\bar{\lambda}(x) - R\bar{\mu}(x) = 0,$$

where $\bar{\lambda}_i(x) = \lambda_i \hat{\lambda}_i(x)$ and $\bar{\mu}_i(x) = \mu_i \hat{\mu}_i(x)$,

(A4) There exists a Lipschitz-continuous function $\hat{a}(x)$ such that

$$\sqrt{n}(\bar{\lambda}^n(\sqrt{n}x) - R\bar{\mu}^n(\sqrt{n}x)) \rightarrow \hat{a}(x)$$

as $n \rightarrow \infty$ uniformly on compact subsets of S , where $\bar{\lambda}_i^n(x) = \lambda_i^n \hat{\lambda}_i^n(x)$ and $\bar{\mu}_i^n(x) = \mu_i^n \hat{\mu}_i^n(x)$.

Then the process M in (2.3) and (2.4) is a K -dimensional Brownian motion with covariance matrix \mathcal{A} which has entries

$$\mathcal{A}_{ii} = \hat{\lambda}_i(0)\lambda_i^3 a_i + \hat{\mu}_i(0)\mu_i^3 s_i(1 - 2p_{ii}) + \sum_{j=1}^K \hat{\mu}_j(0)\mu_j p_{ji}(1 - p_{ji} + p_{ji}\mu_j^2 s_j) \quad \text{for } i \in \mathbb{K},$$

and

$$\mathcal{A}_{ij} = - \left[\hat{\mu}_i(0)\mu_i^3 s_i p_{ij} + \hat{\mu}_j(0)\mu_j^3 s_j p_{ji} + \sum_{k=1}^K \hat{\mu}_k(0)\mu_k p_{ki} p_{kj}(1 - \mu_k^2 s_k) \right] \quad \text{for } 1 \leq i < j \leq K.$$

An application of Theorem 2.4 yields the following result.

Corollary 4.1. *If, in addition to the assumed hypotheses, condition (A1) holds, then the processes $(\hat{Q}^n(nt)/\sqrt{n}, t \geq 0)$ converge in distribution to the process $(X(t), t \geq 0)$ with*

$$X(t) = \Gamma \left(X_0 + \int_0^t \hat{a}(X(s)) ds + \mathcal{A}^{1/2} B(\cdot) \right) (t),$$

where $B(\cdot)$ is a K -dimensional standard Brownian motion.

Remark 4.2. *The conditions on the asymptotics of the arrival and service rates essentially boil down to the assumptions that the following expansions hold: $\bar{\lambda}^n(x) = \bar{\lambda}_1(x/n) + \bar{\lambda}_2(x/\sqrt{n})/\sqrt{n}$ and $\bar{\mu}^n(x) = \bar{\mu}_1(x/n) + \bar{\mu}_2(x/\sqrt{n})/\sqrt{n}$ with suitable functions $\bar{\lambda}_1$, $\bar{\lambda}_2$, $\bar{\mu}_1$, and $\bar{\mu}_2$.*

Remark 4.3. *If, in addition, the assumption of unit rates is made, that is $\hat{\lambda}_j^n(x) = 1$ for $j \in \mathcal{J}$ and $\hat{\mu}_k^n(x) = 1$ for $k \in \mathbb{K}$, then the limit process is a K -dimensional reflected Brownian motion on the positive orthant with infinitesimal drift $\hat{a}(0)$ and covariance matrix \mathcal{A} , and the reflection matrix $R = I - P^T$, as in Theorem 1 of Reiman [18].*

Remark 4.4. *In order to extend applicability, one may consider independent sequences of weakly dependent random variables $\{u_j^i(n), i \geq 1\}$, $\{v_k^i(n), i \geq 1\}$ for $j \in \mathcal{J} \subseteq \mathbb{K}$ and $k \in \mathbb{K}$. Under suitable moment and mixing conditions which imply the invariance principle, cf., e.g., Herrndorf [8], Peligrad [16], Jacod and Shiryaev [9], Corollary 4.1 continues to hold.*

Appendix

Proof of Lemma 2.1. The proof is an adaptation of the one in Puhalskii and Simon [17, Lemma 2.1] and employs the approach of Ethier and Kurtz [6, Theorem 4.1, p.327]. Let

$$\begin{aligned}\theta^n(x) &= 1 + \sum_{i=1}^K (\mu_i^n(x) + \lambda_i^n(x)), \\ \hat{\mu}_i^n(x) &= \frac{\mu_i^n(x)}{\theta^n(x)}, \\ \hat{\lambda}_i^n(x) &= \frac{\lambda_i^n(x)}{\theta^n(x)},\end{aligned}$$

and

$$\tau^n(t) = \inf\{s : \int_0^s \theta^n(Q^n(u)) du > t\}.$$

We note that $\tau^n(t)$ is finite-valued, differentiable, $d\tau^n(t)/dt = 1/\theta^n(Q^n(\tau^n(t)))$ and $\tau^n(t) \rightarrow \infty$ as $t \rightarrow \infty$. One can see that if the process Q^n satisfies a.s. the equations

$$\begin{aligned}Q_i^n(t) &= Q_i^n(0) + N_i^{A,n} \left(\int_0^t \lambda_i^n(Q^n(s)) ds \right) \\ &+ \sum_{j=1}^K \Phi_{ji}^n \left(N_j^{D,n} \left(\int_0^t \mu_j^n(Q^n(s)) 1_{\{Q_j^n(s) > 0\}} ds \right) \right) \\ &- N_i^{D,n} \left(\int_0^t \mu_i^n(Q^n(s)) 1_{\{Q_i^n(s) > 0\}} ds \right), t \geq 0,\end{aligned}\tag{A1}$$

then the process $\hat{Q}^n = (\hat{Q}^n(t), t \geq 0)$ defined by $\hat{Q}^n(t) = Q^n(\tau^n(t))$ satisfies a.s. the equations

$$\begin{aligned}\hat{Q}_i^n(t) &= \hat{Q}_i^n(0) + N_i^{A,n} \left(\int_0^t \hat{\lambda}_i^n(\hat{Q}^n(s)) ds \right) \\ &+ \sum_{j=1}^K \Phi_{ji}^n \left(N_j^{D,n} \left(\int_0^t \hat{\mu}_j^n(\hat{Q}^n(s)) 1_{\{\hat{Q}_j^n(s) > 0\}} ds \right) \right) \\ &- N_i^{D,n} \left(\int_0^t \hat{\mu}_i^n(\hat{Q}^n(s)) 1_{\{\hat{Q}_i^n(s) > 0\}} ds \right), t \geq 0.\end{aligned}\tag{A2}$$

On the other hand, suppose a \mathbb{Z}_+^K -valued process \hat{Q}^n satisfies a.s. (A2) and let

$$\hat{\tau}^n(t) = \inf\{s : \int_0^s \frac{1}{\theta^n(\hat{Q}^n(u))} du > t\}.$$

We show that $\hat{\tau}^n(t)$ is well defined for all t a.s. Since by hypotheses, for a suitable constant L^n , $\theta^n(x) \leq L^n(1+x)$, we have that

$$\int_0^s \frac{1}{\theta^n(\hat{Q}^n(u))} du \geq \frac{1}{L^n} \int_0^s \frac{1}{1 + \sum_{i=1}^K \hat{Q}_i^n(u)} du \geq \frac{1}{L^n} \int_0^s \frac{1}{1 + \sum_{i=1}^K \hat{Q}_i^n(0) + \sum_{i=1}^K N_i^{A,n}(u)} du, \quad (\text{A3})$$

where the latter inequality uses the fact that by (A2)

$$\sum_{i=1}^K \hat{Q}_i^n(t) \leq \sum_{i=1}^K \hat{Q}_i^n(0) + \sum_{i=1}^K N_i^{A,n} \left(\int_0^s \hat{\lambda}_i^n(\hat{Q}_i^n(u)) du \right)$$

and that $\hat{\lambda}_i^n(x) \leq 1$. Since $\limsup_{t \rightarrow \infty} N_i^{A,n}(t)/t < \infty$ a.s., the rightmost integral in (A3) tends to infinity as $t \rightarrow \infty$ a.s., so does the leftmost integral, which proves the claim. In addition, $\hat{\tau}^n(t)$ is differentiable, $d\hat{\tau}^n(t)/dt = \theta^n(\hat{Q}^n(\hat{\tau}^n(t)))$ and $\hat{\tau}^n(t) \rightarrow \infty$ as $t \rightarrow \infty$ a.s. It follows that $Q^n(t) = \hat{Q}^n(\hat{\tau}^n(t))$ satisfies (A1) a.s.

Thus, existence and uniqueness for (A1) holds if and only if existence and uniqueness holds for (A2). The existence and uniqueness for (A2) follows by recursion on the jump times of \hat{Q}^n . In some more detail, we define the processes $\hat{Q}^{n,\ell} = (\hat{Q}^{n,\ell}(t), t \geq 0)$ with $\hat{Q}^{n,\ell}(t) = (\hat{Q}_i^{n,\ell}(t), i = 1, 2, \dots, K)$ by $\hat{Q}_i^{n,0}(t) = \hat{Q}_i^n(0)$ and, for $\ell = 1, 2, \dots$, by

$$\begin{aligned} \hat{Q}_i^{n,\ell}(t) &= \hat{Q}_i^n(0) + N_i^{A,n} \left(\int_0^t \hat{\lambda}_i^n(\hat{Q}^{n,\ell-1}(s)) ds \right) \\ &\quad + \sum_{j=1}^K \Phi_{ji}^n \left(N_j^{D,n} \left(\int_0^t \hat{\mu}_j^n(\hat{Q}^{n,\ell-1}(s)) 1_{\{\hat{Q}_j^{n,\ell-1}(s) > 0\}} ds \right) \right) \\ &\quad - N_i^{D,n} \left(\int_0^t \hat{\mu}_i^n(\hat{Q}^{n,\ell-1}(s)) 1_{\{\hat{Q}_i^{n,\ell-1}(s) > 0\}} ds \right). \end{aligned}$$

Let $\tau^{n,\ell}$ represent the time epoch of the ℓ th jump of $\hat{Q}^{n,\ell}$ with $\tau^{n,0} = 0$. One can see that $\hat{Q}^{n,1}(t) = \hat{Q}^{n,0}(0)$ if $t < \tau^{n,1}$. It follows that $(\hat{Q}^{n,1}(t), t \geq 0)$ and $(\hat{Q}^{n,2}(t), t \geq 0)$ experience the first jump at the same time epoch and the jump size is the same for both processes, so $\tau^{n,1} < \tau^{n,2}$ and $\hat{Q}^{n,1}(t \wedge \tau^{n,1}) = \hat{Q}^{n,2}(t)$ for $t < \tau^{n,2}$. We define $\hat{Q}^n(t) = \hat{Q}^n(0)$ for $t < \tau^{n,1}$ and $\hat{Q}^n(t) = \hat{Q}^{n,1}(t)$ for $\tau^{n,1} \leq t < \tau^{n,2}$. Similarly, for an arbitrary $\ell \in \mathbb{N}$, we obtain that $\tau^{n,\ell} < \tau^{n,\ell+1}$ and $\hat{Q}^{n,\ell}(t \wedge \tau^{n,\ell}) = \hat{Q}^{n,\ell+1}(t)$ for $t < \tau^{n,\ell+1}$. We let $\hat{Q}^n(t) = \hat{Q}^{n,\ell}(t)$ for $\tau^{n,\ell} \leq t < \tau^{n,\ell+1}$. The process \hat{Q}^n is defined consistently for $t \in \cup_{\ell=1}^\infty [\tau^{n,\ell-1}, \tau^{n,\ell})$. If $\tau^{n,\ell+1} = \infty$ for some ℓ , then we let $\hat{Q}^n(t) = \hat{Q}^{n,\ell}(t)$ for all $t \geq \tau^{n,\ell}$.

Suppose that $\tau^{n,\ell} < \infty$ for all ℓ . Then $\hat{Q}^n(t)$ has been defined for all $t < \tau^{n,\infty} = \lim_{\ell \rightarrow \infty} \tau^{n,\ell}$ and satisfies (A2) for these values of t . We now show that $\tau^{n,\infty} = \infty$. The set of the time epochs of the jumps of \hat{Q}^n is a subset of the set of the time epochs of the jumps

of the process $\tilde{Q}^n = (\tilde{Q}^n(t), t \geq 0)$, where

$$\begin{aligned} \tilde{Q}^n(t) = & \sum_{i=1}^K \left(N_i^{A,n} \left(\int_0^t \hat{\lambda}_i^n(\hat{Q}^{n,\ell-1}(s)) ds \right) + \sum_{j=1}^K \Phi_{ji}^n \left(N_j^{D,n} \left(\int_0^t \hat{\mu}_j^n(\hat{Q}^{n,\ell-1}(s)) 1_{\{\hat{Q}_j^{n,\ell-1}(s) > 0\}} ds \right) \right) \right. \\ & \left. + N_i^{D,n} \left(\int_0^t \hat{\mu}_i^n(\hat{Q}^{n,\ell-1}(s)) 1_{\{\hat{Q}_i^{n,\ell-1}(s) > 0\}} ds \right) \right). \end{aligned}$$

Since the process \hat{Q}^n has infinitely many jumps, so does the process \tilde{Q}^n . Since $\hat{\lambda}_i^n(x) \leq 1$, $\mu_i^n(x) \leq 1$ and $\Phi_{ji}^n(m_1) - \Phi_{ji}^n(m_2) \leq m_1 - m_2$ for $m_1 \geq m_2$, the lengths of time between the jumps of \tilde{Q}^n are not less than the lengths of time between the corresponding jumps of the process $(\sum_{i=1}^K N_i^{A,n}(t) + \sum_{i=1}^K N_i^{D,n}(t), t \geq 0)$. The process $(\sum_{i=1}^K N_i^{A,n}(t) + \sum_{i=1}^K N_i^{D,n}(t), t \geq 0)$ having infinitely many jumps, the time epochs of the jumps of $(\sum_{i=1}^K N_i^{A,n}(t) + \sum_{i=1}^K N_i^{D,n}(t), t \geq 0)$ tend to infinity as the jump numbers tend to infinity. Thus, $\tau^{n,\infty} = \infty$ a.s.

The provided construction shows that Q^n is a suitably measurable function of $N^{A,n}$, $N^{D,n}$, and Φ^n , so it is a strong solution. We have proved the existence of a strong solution to (A2). A similar argument establishes uniqueness. \blacksquare

Acknowledgments

The research work of Chihoon Lee is supported in part by the Army Research Office (W911NF-12-1-0494).

References

- [1] R. Bekker. *Queues with state-dependent rates*. PhD thesis, Technische Universiteit Eindhoven, Eindhoven, 2005.
- [2] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [3] W. K. Chan and L. W. Schruben. Mathematical programming representations for state-dependent queues. In S.J. Mason, R.R. Hill, L. Mönch, O. Rose, T. Jefferson, and J.W. Fowler, editors, *Proceedings of the 2008 Winter Simulation Conference*, pages 495–501. IEEE, 2008.
- [4] P. Dupuis and H. Ishii. On Lipschitz continuity of the solution mapping to the Skorohod problem, with applications. *Stochastics*, 35:31–62, 1991.
- [5] P. Dupuis and H. Ishii. SDEs with oblique reflection on nonsmooth domains. *Ann. Probab.*, 21(1):554–580, 1993. Correction: *Ann. Probab.*, 36(5):1992–1997, 2008.

- [6] S. N. Ethier and T. G. Kurtz. *Markov Processes. Characterization and Convergence*. Wiley, 1986.
- [7] J. M. Harrison and M. I. Reiman. Reflected Brownian motion on an orthant. *Ann. Probab.*, 9(2):302–308, 1981.
- [8] N. Herrndorf. A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.*, 12(1):141–153, 1984.
- [9] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
- [10] E. V. Krichagina. Asymptotic analysis of queueing networks (martingale approach). *Stochastics Stochastics Rep.*, 40(1-2):43–76, 1992.
- [11] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes Appl.*, 6(3):223–240, 1977/78.
- [12] H. J. Kushner. *Heavy traffic analysis of controlled queueing and communication networks*, volume 47 of *Applications of Mathematics*. Springer-Verlag, New York, 2001.
- [13] A. Mandelbaum, W. A. Massey, and M. Reiman. Strong approximations for Markovian service networks. *Queueing Syst.*, 30:149–201, 1998.
- [14] A. Mandelbaum and G. Pats. State-dependent stochastic networks. I. Approximations and applications with continuous diffusion limits. *Ann. Appl. Probab.*, 8(2):569–646, 1998.
- [15] G. Pang, R. Talreja, and W. Whitt. Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.*, 4:193–267, 2007.
- [16] M. Peligrad. Invariance principles under weak dependence. *J. Multivariate Anal.*, 19(2):299–310, 1986.
- [17] A. Puhalskii and B. Simon. Discrete evolutionary birth-death processes and their large population limits. *Stochastic Models*, 28(3):388–412, 2012.
- [18] M. I. Reiman. Open queueing networks in heavy traffic. *Mathematics of Operations Research*, 9(3):441–458, 1984.
- [19] W. Whitt. Queues with service times and interarrival times depending linearly and randomly upon waiting times. *Queueing Systems Theory Appl.*, 6(4):335–351, 1990.
- [20] K. Yamada. Diffusion approximation for open state-dependent queueing networks in the heavy traffic situation. *Ann. Appl. Probab.*, 5(4):958–982, 1995.